U.S. DEPARTMENT OF COMMERCE
National Technical Information Service
PB-273 513

Guide to Preselection of Training Samples and Ground Truth Collection

Lockheed Electronics Co, Inc, Las Vegas, Nev Remote Sensing Lab

Prepared for

Environmental Monitoring and Support Lab, Las Vegas, Nev

Sep 77

EPA

PB 273 513

GUIDE TO PRESELECTION OF TRAINING SAMPLES AND GROUND TRUTH COLLECTION

Interagency
Energy-Environment
Research and Development
Program Report

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U. S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

RESEARCH REPORTING SERIES

Research reports of the Office of Research and Development, U.S. Environmental Protection Agency, have been grouped into nine series. These nine broad categories were established to facilitate further development and application of environmental technology. Elimination of traditional grouping was consciously planned to foster technology transfer and a maximum interface in related fields. The nine series are:

1. Environmental Health Effects Research
2. Environmental Protection Technology
3. Ecological Research
4. Environmental Monitoring
5. Socioeconomic Environmental Studies
6. Scientific and Technical Assessment Reports (STAR)
7. Interagency Energy-Environment Research and Development
8. "Special" Reports
9. Miscellaneous Reports

This report has been assigned to the INTERAGENCY ENERGY-ENVIRONMENT RESEARCH AND DEVELOPMENT series. Reports in this series result from the effort funded under the 17-agency Federal Energy/Environment Research and Development Program. These studies relate to EPA's mission to protect the public health and welfare from adverse effects of pollutants associated with energy systems. The goal of the Program is to assure the rapid development of domestic energy supplies in an environmentally-compatible manner by providing the necessary environmental data and control technology. Investigations include analyses of the transport of energy-related pollutants and their health and ecological effects; assessments of, and development of, control technologies for energy systems; and integrated assessments of a wide range of energy-related environmental issues.

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

TECHNICAL REPORT DATA

1. REPORT NO.	2.	3. RECIPIENT'S ACCESSION NO.
EPA-600/7-77-100		PB 273513

4. TITLE AND SUBTITLE	5. REPORT DATE
GUIDE TO PRESELECTION OF TRAINING SAMPLES AND GROUND TRUTH COLLECTION	September 1977
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
Charles E. Tanner	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Lockheed Electronics Company, Inc. Remote Sensing Laboratory Las Vegas, Nevada 89114	10. PROGRAM ELEMENT NO. EHE 625
	11. CONTRACT/GRANT NO. EPA 68-03-2153
12. SPONSORING AGENCY NAME AND ADDRESS U.S. Environmental Protection Agency/Las Vegas, NV Office of Research and Development Environmental Monitoring and Support Laboratory Las Vegas, Nevada 89114	13. TYPE OF REPORT AND PERIOD COVERED
	14. SPONSORING AGENCY CODE EPA/600/07

15. SUPPLEMENTARY NOTES

16. ABSTRACT

This report was prepared to provide the novice data processing analyst and field personnel with the tools and basic concepts used in the processing of multispectral scanner data via an interactive or conventional processing system.

Included in the guide is an explanation of the need for the collection of accurate/inexpensive "ground truth" and brief descriptions of the various ecosystems that will be encountered in this study. Also, a detailed list of the actual parameters that should be included in a well-designed ground truth form are provided.

Sampling schemes from Landsat and aircraft multispectral scanner data are also discussed at length along with procedures and recommendations for selecting training samples from photography for use in automatic data processing.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Aerial Photography Land Use Photographic Reconnaissance Photo Interpretation Space Born Photography Stereo Photography	Automatic Data Processing Multispectral Scanner Ground Truth Field Observations	10B 14E

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS (*This Report*)	21. NO. OF PAGES
Release to Public	Unclassified	32
	20. SECURITY CLASS (*This page*)	22. PRICE
	Unclassified	PC A03 - A01

EPA Form 2220-1 (9-73)

☆ U.S. GOVERNMENT PRINTING OFFICE: 1977-785-007/1007 9-1

EPA-600/7-77-100
September 1977

GUIDE TO PRESELECTION OF TRAINING SAMPLES AND GROUND TRUTH COLLECTION

by

Charles E. Tanner
Lockheed Electronics Company, Inc.
Las Vegas, Nevada 89114

Contract 68-03-2153

Project Officer

Robert W. Landers
Remote Sensing Division
Environmental Monitoring and Support Laboratory
Las Vegas, Nevada 89114

ENVIRONMENTAL MONITORING AND SUPPORT LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
LAS VEGAS, NEVADA 89114

DISCLAIMER

This guide has been reviewed by the Environmental Monitoring and Support Laboratory/Las Vegas, Nevada, U.S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

FOREWORD

Protection of the environment requires effective regulatory actions which are based on sound technical and scientific information. This information must include the quantitative description and linking of pollutant sources, transport mechanisms, interactions, and resulting effects on man and his environment. Because of the complexities involved, assessment of specific pollutants in the environment requires a total systems approach which transcends the media of air, water, and land. The environmental Monitoring and Support Laboratory-Las Vegas contributes to the formation and enhancement of a sound integrated monitoring data base through multidisciplinary, multimedia programs designed to:

- develop and optimize systems and strategies for monitoring pollutants and their impact on the environment
- demonstrate new monitoring systems and technologies by applying them to fulfill special monitoring needs of the Agency's operating programs

This report describes and outlines procedures for the preselection of training samples used in computer processing of multispectral scanner data. These data are then used to access reclamation efforts and monitor changes on active strip mines in the Western United States.

George B. Morgan
Director
Environmental Monitoring and Support Laboratory
Las Vegas, Nevada

CONTENTS

Foreword iii
Figures vi
Tables vi
Acknowledgment vii

1. Introduction 1
2. Summary 3
3. Conclusions and Recommendations 4
4. Geography 5
    Geography and History of the Area 5
    Mid-latitude Grasslands 5
    Desert Vegetation 6
    Mountain Vegetation 6
    Vegetative Groupings 7
5. Sampling Scheme 9
    Landsat Scheme 9
    Aircraft Scheme 14
6. Training Sample Selection 17
7. Ground Truth 19
    Information Requirements 19
    Field Operations 19
    Data Collection Equipment and Materials 20
        The Ground Truth Form 20
        Additional Equipment 20

References 23

Preceding page blank

FIGURES

Number		Page
1	One of many possible line-scanning methods utilized in airborne infrared sensing	10
2	Black and white Landsat image of eastern Montana	12
3	Low-cost Data System	13
4	Image processing system of the Data Analysis Station	15

TABLES

Number		Page
1	Simulated Classification Hierarchy	11
2	Modified ground truth form	21

ACKNOWLEDGMENT

The scale and scope of this report were made possible through the cooperation of the U.S. Environmental Protection Agency and the National Aeronautics and Space Administration/Earth Resources Laboratory (NASA/ERL) in Slidell, Louisiana.

Sincere thanks are extended to Mr. Larry Erickson, Lockheed Electronics Company at NASA/ERL, for allowing us to cite sections of his unpublished manuscript and also for acting as major advisor on this report.

Appreciation is extended to Mr. Sidney Whitley for providing the necessary documentation for generating the photography of the Varian-75 system and peripheral hardware.

SECTION 1

INTRODUCTION

Quite often data processing analysts are forced to process multispectral scanner data with a minimum of background information and personal knowledge of the scene under investigation. This situation is compounded by the collection of inadequate data about the site. Under ideal situations a data sampling scheme should be developed by the project statistician with inputs from the data processing analyst.

The classification hierarchy (a categorical ranking of the natural and/or manmade features for use in processing digital multispectral scanner data) and ground truth form (a ground truth form is merely an out-of-doors extension of the classification hierarchy) should be developed by the person(s) responsible for the data processing in conjunction with the field mensuration leader. This would assure the collection of adequate ground data.

Obviously there are certain things that cannot be done in the field because of time and resource constraints. On the other hand, there are certain operations that can only be accomplished by a knowledgeable field person with virtually little expenditure of energy, time, and/or resources. These points must be discussed at meetings that are open to free discourse of ideas and suggestions. An explanation of the data processing objectives should also be presented to the field personnel to make them cognizant that a thorough job on their part will greatly enhance the analyst's chances of generating an accurate classification map and that they are also responsible for the results and any conclusions drawn from such analysis.[1]

Field ground truth operations are usually expensive and time consuming and rarely have been done properly. New efficient methods must be developed to make it much more cost-effective. Hopefully, statistical sampling methods will be developed in order to reduce the total number of ground truth samples. In essence, ground truth data must be collected by statistically sound methods and must be efficiently collected and sufficiently accurate for the given application and objectives.

This report addresses the problem of how to go about collecting ground truth for the purpose of processing digital data. Also, this report outlines some of the procedures used by the author in the actual data reduction phase of automatic data processing. Because of the immediate need of such a report by personnel engaged in ground truth collection at the Environmental Monitoring and Support Laboratory in Las Vegas, Nevada, no attempt was made to familiarize the reader with the basic principles of pattern recognition analysis.

However, those who are interested in pursuing this type of analysis may refer to Landgrebe's "Systems Approach to the Use of Remote Sensing" published by Purdue University in 1971.

SECTION 2

SUMMARY

In summary, pattern recognition using the multispectral approach has been described as an analysis procedure that has proved useful in coping with the vast amount of digital data being collected daily by conventional aircraft and space satellites (Landsat). This type of analysis is heavily dependent upon accurate ancillary data, e.g., topographic maps, vegetation maps, geologic survey, land-use maps and ground truth. Current ground truth that has been accurately and inexpensively gathered is a "dream come true" for the automatic data processing analyst.

This guide addresses the criteria and procedures that must be used in the gathering of ground truth data for use in processing digital remote sensing data. Regardless of the choice of pattern recognition analysis, this report should have application to the ground truthing activities.

SECTION 3

CONCLUSIONS AND RECOMMENDATIONS

Ground truth operations are usually expensive and time consuming and are rarely performed properly. The pattern recognition community in the world of remote sensing is indeed dependent upon ground truth data and can do very little, if anything, without this vital information. Work must continue to develop new efficient methods to make ground truth collection more cost-effective. Also work must continue on the development of statistical sampling methods that will reduce the total number of ground truth samples.

The procedures outlined in this guide were based on specific agency constraints, systems designs and manpower availability; therefore, it is recommended that modifications be made to the procedure to fit the needs of the analyst. It is also recommended that an explanation of the data processing objectives be presented to the field mensuration personnel to make them aware of the need for the collection of accurate ground truth. Too, it is recommended that preselection of training samples be performed as a joint venture between the analyst and the field team. Such an effort would foster cooperation as well as providing a learning experience for the novice field personnel.

SECTION 4

GEOGRAPHY OF THE AREAS

Before a classification hierarchy or sampling scheme can be developed, one must obtain basic information concerning the area. The following information is intended to illustrate the broad ranges of problems associated with developing inventory techniques for strip mining operations via remote sensing techniques in the west and Northern Great Plains areas of the United States.

The Western Energy Project/Strip Mining (WEP/SM) is a rather unusual undertaking. It is the analyst's intent to map active mining features, revegetative types within the mined area and natural vegetation and also determine the condition and density classes associated with each vegetative species via automatic data processing techniques.

GEOGRAPHY AND HISTORY OF THE AREA

The Northern Great Plains are expansive prairie lands bounded in the west by the Rocky Mountains in Montana and Wyoming. The whole Northern Great Plains Coal Province includes all coal in western North Dakota, coal occurring in the Missouri River drainage and east of the Rocky Mountains of Montana and Wyoming, coal in western South Dakota, coal in the Denver Basin of Colorado and Raton Mesa of Colorado and eastern New Mexico. A good deal of the Federal Coal Reserve lies in this province.

The geologic formation containing these coal deposits consists of from 1,700 to 3,200 feet of sandstone, shale and coal. In general, coalbeds are thickest in Wyoming. Some important coalbeds are the Badges and School seams in Glenrock, the Monarch seam near Sheridan, the Healey bed near Lake DeSmet, and the famous Wyodak bed near Gillette.[2] These coal seams in this region were formed from thick and extensive accumulations of biological, principally vegetative, matter buried through geologic time. Development of the thick coal seams of the west required very large flooded areas (swamps) which slowly subsided while growth of vegetation was optimal.

MID-LATITUDE GRASSLANDS

In the text written by Glen Trewartha, 1961 [3], page 322, it states that, "In the midcontinent grasslands of the interior United States east of the Rockies, two, or sometimes three, subdivisions are recognized. The tall-grass prairie, or true prairie, which has been replaced by farmland, used to occupy the more humid eastern parts, and the short-grass steppe, or

plains, still exists in many of the drier western parts. Between them is a transition zone called the mixed-grass prairie, which consists both of midgrasses, two to four feet tall, and of shorter grasses, which together form an upper and a lower story. There are assumptions that the two-story mixed-grass prairie originally also prevailed on the plains area farther west, and that the midgrass species were largely killed by overgrazing. West of the Rockies extensive vegetation areas of the original short bunchgrass still exist, mainly in Washington, Oregon, Idaho and California."(3)

DESERT VEGETATION

In the same text (3) page 323, Trewartha said, "Desert type vegetation pose still another problem. Some soilless, windswept, rocky deserts and some areas of moving sand dunes, may be devoid of all vegetation and these are the exception. Most desert regions have some plant life, although almost invaribly it is sparse. Desert plants are of several types and each has a different way of surmounting the handicaps of its arid environment."

"Desert shrub is the most widespread form of arid-land vegetation. The deciduous species make use of leaf shedding in order to withstand drought. The evergreen varieties have protective structures such as small, thick and leathery leaves with shiny, waxy surfaces. The two most widespread desert shrubs are sagebrush and creosote bush, the second of which grows chiefly in the hotter and more arid southwest."

"Arid-land vegetation also includes leafless, thorny succulents such as cacti, certain salt-tolerant plants, and a variety of short-lived transients. The so-called transients, usually small in size, include many flowering annuals, but also grasses and tuberous plants."

MOUNTAIN VEGETATION

According to Trewartha (3) "Mountain areas are characterized by an unusual variety of local environments existing in close juxtaposition. On the lower slopes of highlands, the vegetation cover may resemble that of the surrounding lowlands. With increasing altitude, temperature decreases rapidly, while solar energy, rainfall and wind speeds increase. Change of vegetation fall into a rough vertical zonation of plant life, but with many interrupting local variations."

"Conifers, with their strong tolerance for the vicissitudes of mountain climates and soils, comprise the largest element of mountain forests in middle latitudes and they are also found at high elevations. Within the coniferous zone, pines usually predominate at lower elevation, but fir and spruce take over near the upper climatic limits of forest. Above these limits, which vary in altitude, trees will not grow because of low temperatures, a short growing season, diurnal freeze and thaw, strong winds and thin soils that alternate between saturation and aridity."(3)

VEGETATIVE GROUPINGS

Paul Packer in his technical report entitled "Rehabilitation Potentials and Limitations of Surface-Mined Land in the Northern Great Plains"[2] identified 16 broad, but recognizable, vegetation types. Nine of these 16 vegetation types occur on surface mineable areas. These vegetation types and an explanation as to their range and suitability as rehabilitation types are also given below:

Floodplain -- This vegetation type occurs in the bottom lands and banks of major rivers and on broad floodplain terraces that have alluvial soils. It occupies sites, some highly saline, that have high water tables. The floodplain type is characterized predominately by hardwood tree and shrub species that are poorly suitable and poorly available for rehabilitation.

Badlands -- This vegetation type occupies breaks along rivers and streams and steep south slopes of exposed shales, sandstones and clays. Dominant plant species are arid-land shrubs and grasses associated locally with scabby ponderosa pine forests. The species that characterize this type are poorly available for rehabilitation.

Short-Grass Prairie -- This vegetation type occupies dry prairies on shallow soils in southeastern Montana and northeastern Wyoming. Dominant species are blue grama grass, western wheatgrass, and various needlegrasses. The species that characterize this type have moderately poor suitability and fair availability for rehabilitation.

Mid-Short Grass Prairie -- This type occupies rolling prairies on loam to clay-loam soils in eastern Montana. It is characterized by western wheatgrass, needle and thread grass and blue grama grass. These species have moderately poor suitability and fair availability for rehabilitation.

Mid-Grass Prairie -- This type, which occurs on loamy soils in extreme eastern Montana, southwestern North Dakota, and northwestern South Dakota, contains no dominant short grass. Principal species are needlegrasses, wheatgrasses, and blue stem grasses. Most species that comprise this type are fairly suitable and have good availability for rehabilitation.

Grassland-Sagebush -- This vegetation type occurs on open grassland of mid and short grasses, with scattered sagebrush, and occurs on silty clay-loam soils in southeastern Montana and northeastern Wyoming. Most of the species that comprise this type have good suitability and moderately good availability for rehabilitation.

Sagebrush-Steppe -- This type is dominated by sagebrush in open grassland containing wheatgrasses, needlegrasses and threadgrasses on silt to silty clay-loam soils. It occurs chiefly in north-

eastern Wyoming and most of the major species have moderately good suitability, but are relatively unavailable for rehabilitation.

Mid-Tall Grass Prairie -- This vegetation type occurs on gently rolling prairie northeast of the Missouri River in North Dakota, and is characterized by wheatgrasses, big and little blue stem grasses, and needlegrasses on loam soils of glacial till origin. Nearly all species in this type have good suitability and availability for rehabilitation.

Ponderosa Pine Forest -- This vegetation type occurs mainly in eastern Montana and northeastern Wyoming on uplands, ridges and north slopes that have shallow loam soils. Prominent species are ponderosa pine, snowberry, blue grasses, fescues, and June grass. These species are only fairly suitable but have good availability for rehabilitation.

These types have been modified and placed in the Western Energy Project/ Strip Mining classification hierarchy for use in the actual data reduction phase of the study.

SECTION 5

SAMPLING SCHEME

Before a scene (any segment of digital data) can be classified distinct signatures must be developed for each class in the hierarchy. This process of developing signatures poses the problem of determining the number of training fields that must be selected for such an undertaking. Too many training fields would require an excessive expenditure of resources in the field as well as increasing processing time in the Laboratory. Conversely, an insufficient number of training fields that are poorly distributed over the imagery could be the cause of the generation of a classification run with little, if any, confidence attached to the results. Also, scanner anomalies will require that training fields be selected at nadir and to the left and right of this point as well.

After all is said and done, it must be realized that there is, at this time, no way to systematically establish training fields for a given number of classes over a large area simply because classes will occur as nature dictates, e.g., certain tree species can only be found where the soil types, elevation and water regime meet their requirements. Training fields have to be established wherever the class of interest occurs along the entire length of the flight line or across the scanner's full field of view (Figure 1).

LANDSAT SCHEME

Assume that, based on the information in the previous section, Table 1 has been developed as the initial step in attempting to use Landsat data to classify native vegetation in eastern Montana. This region is characterized by a short summer season of scanty precipitation and low humidities with a high percentage of clear, warm, sunny days and winters with heavy snowfall and fairly low temperatures. Annual precipitation ranges from 28 to 60 inches with the fall season receiving the largest percentage of rainfall. Growing season (May through August) precipitation varies from 5 to somewhat more than 7 inches. (4) Vegetation types are in the mid-grass and mid-short grass prairie varieties as described earlier.

A reduction of a Landsat image of this area may be seen in Figure 2. This imagery is formed by merging the four tapes that are required to complete a scene into a single 9-track computer compatible tape. Following this merging step, the tape is viewed and filmed at the Data Analysis Station (Figure 3). Finally, the 9" x 9" film format is processed according to standard operational procedures to produce transparencies or hard copies. Training samples should be selected from this imagery in conjunction with

SCANNER GEOMETRY

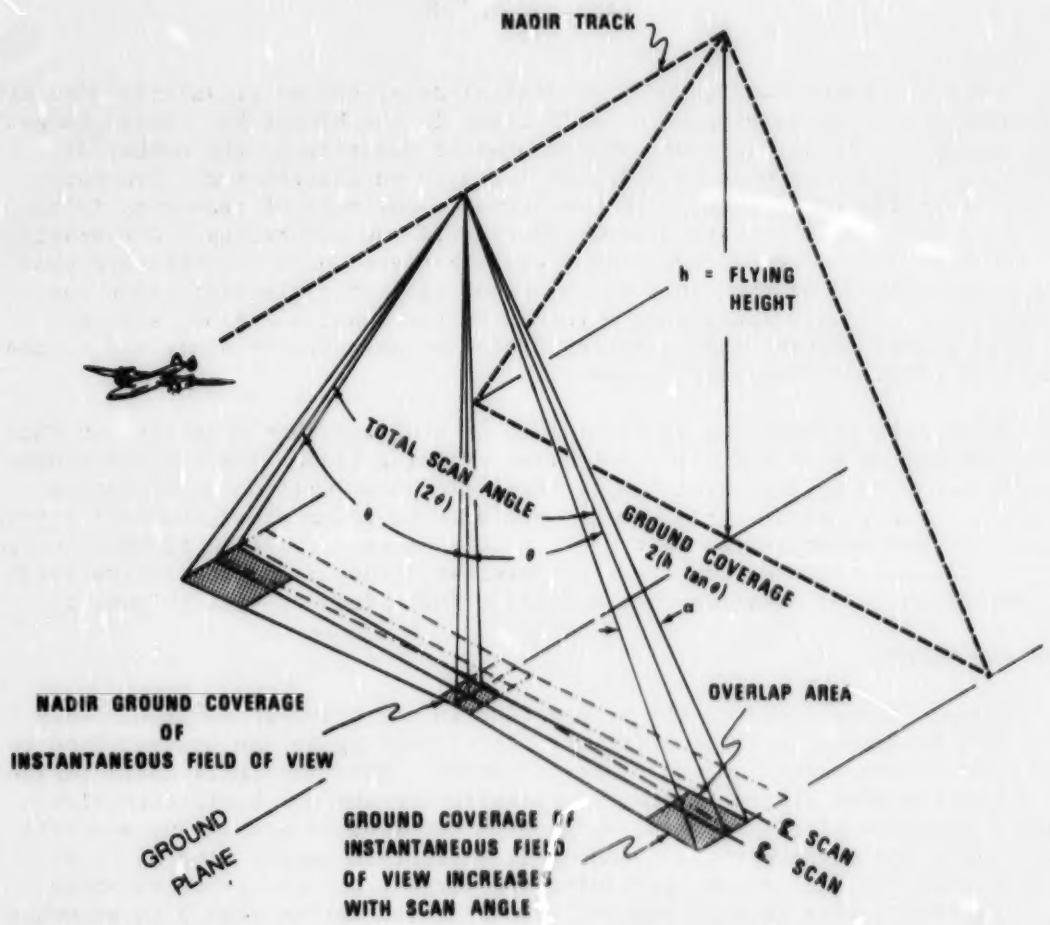

Figure 1. One of many possible line-scanning methods utilized in airborne infrared sensing. A rotating mirror in the multispectral scanner scans the terrain perpendicular to the line of flight.

TABLE 1. SIMULATED CLASSIFICATION HIERARCHY.

LEVEL I	LEVEL II	LEVEL III	LEVEL IV	LEVEL V
Agriculture	Cultivated Fields			
	Fallow Fields			
Forestland	Pine	White Pine	0-20% Canopy Density	Seedling/Saplings
		Lodgepole Pine	25-50% Canopy Density	Pole Timber
		Ponderosa Pine	50-75% Canopy Density	Immature Sawtimber
			75-> % Canopy Density	Mature Sawtimber
	Hardwood			
Grasslands	Short-Grass Prairie	Blue Grama	0-25% Ground Density	Seedlings
	Mid-Grass/Mid-Short Grass	Wheatgrass	25-50% Ground Density	Immature
	Grassland - Sagebrush	Needlegrass	50-75% Ground Density	Mature
			75-> % Ground Density	Mixed

Figure 2. NASA black and white Landsat imagery of eastern Montana.

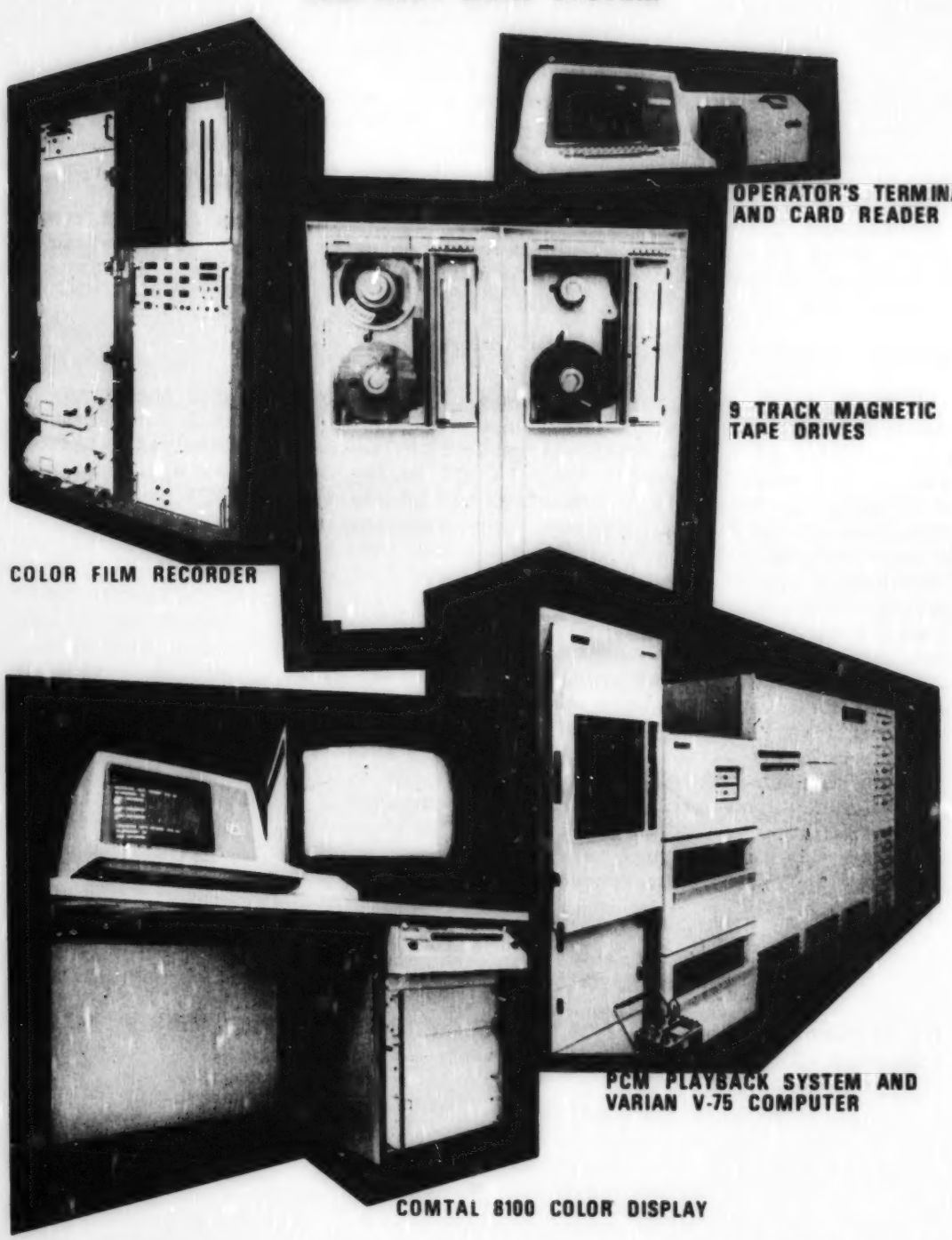
LOW-COST DATA SYSTEM
OPERATOR'S TERMINAL AND CARD READER
9 TRACK MAGNETIC TAPE DRIVES
COLOR FILM RECORDER
PCM PLAYBACK SYSTEM AND VARIAN V-75 COMPUTER
COMTAL 8100 COLOR DISPLAY

Figure 3. The Low-Cost Data Analysis Station adapted from NASA/ERL report number 157.

low altitude color-infrared photography of the same area. As a general practice, at least three training samples per class per tape should be selected, outlined and coded on clear or frosted overlay material. Based on the classification hierarchy (Table 1) it is necessary to select at least 12 training samples per specie per tape. This means that 36 training samples for the general class "grassland" would have to be selected. Three training samples per class per tape is by no means the maximum nor the minimum number required for sampling purposes. In areas of intensive land use considerably more samples may be necessary to accurately identify the class because of the varying nature of land use practices between individual land owners.

Because of the difference in line/pixel (instantaneous field of view) count, which is a function of altitude and detector size, a new sampling scheme must be devised to facilitate processing of the aircraft multispectral scanner data.

AIRCRAFT SCHEME

To begin remember the constraints of the system on which the multispectral scanner data are to be processed. For all practical purposes we will confine all of our calculations to the low-cost Data Analysis Station (DAS). The image processing system (IPS) on the DAS (Figure 4) is capable of displaying 508 pixels of the total 628 pixels from the EPA scanner and 508 lines of the data at one time. The sampling strategy for aircraft scanner data will be based on these figures and the assumption that all data have been obtained at an altitude of 12,000 feet with a multispectral scanner having a 2.5 mrad rad. detector size. Also, a 6 x 6 pixel (slightly less than 1 acre of 0.301 hectare) cursor will be used as a standard sample size. Data obtained at the aforementioned altitude with the 2.5 mrad spot size will have a pixel size of approximately 30 feet (9.04 meters) on a side at nadir. Due to scanner sweep angle and instability of the aircraft, pixel sizes will vary from both sides of nadir. This problem can, of course, be corrected by software programs currently available on the system. A one percent sample of 258,000 pixels (508 pixels x 508 pixels of data) or 5,160 acres (2085 hectares) is sufficient to demonstrate the design potential of the automatic processing unit. This is what automatic data processing is all about; performing an accurate inventory with less expenditure of time, money and natural energy sources than conventional mthods. A one percent sample would be:

Total number pixels 258,064 x 0.01 sample rate =
2580.64 pixels (total number of samples)

The standard sample rate is 36 pixels per training field (roughly 3/4 of an acre or 0.03 hectare), therefore:

2580.64 ÷ 36 = 71 total training samples/scene

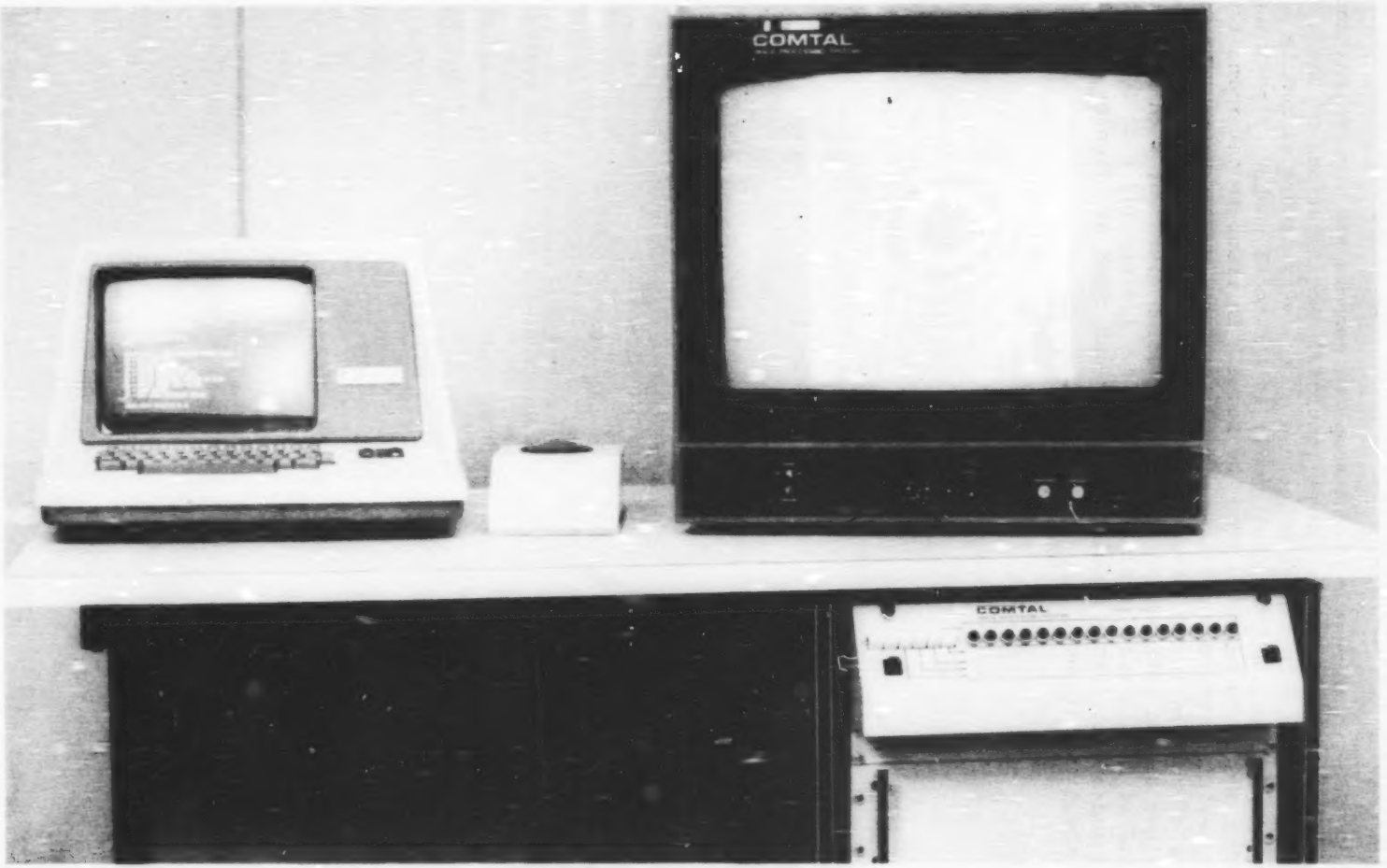

Figure 4. Image Processing System (IPS) of the Data Analysis Station.

It is important to remember that sampling schemes are theoretical and serve as an initial step in most investigations. Modifications of the schemes are and should be expected because of the very nature of remote sensing targets. This derived figure (71) is the total number of training samples that should be selected from the scene. In order to determine the number of training samples per class divide the total number of samples by the number of classes in the hierarchy (in this example the number is 12). This sampling scheme is designed for classification hierarchy with 5 or more classes and/or subclasses. Training samples for hierarchies with less than 5 classes and/or subclasses should be located and coded as often as possible and over the entire scene.

SECTION 6

TRAINING SAMPLE SELECTION

The previous material provided an analyst with a number to work with: a starting point. It is now the responsibility of the photo interpreter to locate areas on the ground, via photographs or transparencies, that are in the classification hierarchy.

Aircraft color, color-infrared, or black and white imagery should be used in the selection of training fields; the optimum condition being that the aerial photography is obtained simultaneously with the aircraft multispectral scanner coverage (all flightlines and other flight parameters should be calculated on scanner specifications since it is the primary sensor). Selecting training fields in this manner, prior to field observations, is at best a guess. The ground training samples may exhibit conditions considerably different from those prognosticated by the image interpreter. There are several criteria that must be satisfied before the training field is accepted, its spectral characteristics examined and ultimately merged to form a single representative spectral signature by which all other areas of the same category or classification may be recognized. The first and most important criterion is that the training field be homogeneous and second its material must be uniformly distributed. This is, of course, determined by the spatial attributes of the field in question. Obviously, homogeneity and uniformity may be affected by a variety of natural or anomalous influences that can occur in a brief period of time. Size is the third criterion used in selecting training fields. Sample size can be measured directly from the photography when the scane of imagery and/or ground distance is known. However, subsequent field visits may reveal dimensional changes, caused by manmade factors or natural phenomena invalidating the sample(s).

In the selection of training samples, the image analyst should attempt to distribute them evenly throughout the geographic area covered by the scanner. A relatively even distribution of samples will in most cases:

1. Account for variations in ground cover conditions over a fairly large geographic area.

2. Force field teams to observe a larger geographic area in which additional samples may be selected to supplement existing ones or replace invalidated samples. In the process of traveling between preselected training samples, field personnel frequently

observe features or conditions unique or common enough to include as additional training samples; by covering a considerable geographic area in the field, the potential for additional data and representative samples is increased immeasurably.

Whether identified by photo interpretation or field observation, every training sample must be precisely located. There should never be any question as to the exact location of a training sample. In fact, whenever possible, training fields should be located so that geographic features can help in locating these samples on the photography, in the field and on the screen of a cathode ray tube. These training samples may be delineated on translucent overlays and annotated with preestablished coding for the various categories in the hierarchy. Fiducial marks and distinctive geographic features such as roadways, pipelines, water bodies, etc., should be traced for assistance in overlay orientation.

The point of all this is that field personnel should never expect to find perfect training samples at every location indicated on the photography. Because of this, programs must be planned with sufficient flexibility to contend with these problems and, yet, gather the quality and quantity of data required to develop an accurate and useful classification.

In addition to surface conditions such as size, homogeneity, and uniformity, anomalies in scanner data may prevent using one or more training samples. It must be recognized that no single sample can be expected to adequately represent every other area of similar surface material within the scanner field of view.

Training sample selection is by no means infallible. When visited in the field, even the most likely training samples may fall short of expectations. Field verification of any group of preselected samples may result in as much as 20% attrition rate. Personnel in the field are in a much better position than the laboratory photo interpreter to evaluate the qualifications of a given site as a potential training sample. All samples identified in this manner should be plotted and coded on maps, imagery and overlays with the same exacting attitude and considerations as in the preselection process.

SECTION 7

GROUND TRUTH

Surface investigations provide indispensable information for developing a valid ground cover classification, whether the remote sensing system employs traditional image analysis or automated data processing techniques.

The term "ground truth" used in remote sensing communities is actually a misnomer. Ground observations are really only valid when they are at the time of overflight. Traditionally, ground truth is acquired a week or two after the overflight and is therefore, subject to speculation. Because of its wide acceptance and usage by the remote sensing community we will continue to refer to these visits (regardless of time) as ground truth.

INFORMATION REQUIREMENTS

A comprehensive land use project must accurately identify as many land uses or ground cover types/conditions as the state of the art allows.

For this to occur, determine what conditions (sun angle, climatic or atmospheric) will have the most significant influence on developing a valid classification for a given project, in a given geographic area at a particular time of the year.

Also, it must be remembered that the Earth's surface is being studied and that although the major physiographic features may seem to remain statis, the ground cover, whether manmade or natural, are extremely dynamic. Since the individual or combined influences of the conditions may vary considerably from one ecosystem to the next, it is vitally important that the area selected as the focus of surface investigations be representative of the entire geographic area to be classified.

FIELD OPERATIONS

In the actual field operations, it is important that the investigators be reasonably well versed in geography, botany, forestry and agronomy, and have basic photo interpretation skills. Each person of a ground truth team should have at least a good understanding of how the remote sensor will respond to the various surface conditions being identified. Field work must be carried out efficiently as well.

DATA COLLECTION EQUIPMENT AND MATERIALS

Regardless of attempted simplicity, ground truth activities are often encumbered by numerous checklists, forms, maps, photographs and reference material that may rapidly become a confusing mass of paper. To minimize this possibility, all vital materials should be filed in an envelope or folder large enough to contain the following items:

Ground truth forms
Training sample check list
Instructions
Aerial photography (paper prints and/or transparencies)
Maps
Photographic logs (if required)

The Ground Truth Forms

A well researched, well human-engineered ground truth form is essential in the automatic data processing area of remote sensing. The data that are obtained on the ground and transferred via the ground truth form serves as the base on which decisions that affect data processing procedures are made.

A site observation form should be easy to complete and can serve as a motivating force to the observer to note and assess those features that appear to be anomalies. The form should be straightforward and designed to eliminate excessive writing which could eventually cause the observer to become sloppy and careless. In constructing the form, attempt to avoid redundant entries and try to maintain a logical ordering of the data. Keep the form to a maximum of two pages (reverse printing is acceptable) and concentrate on those caegories that will satisfy your hierarchial requirements. Not only will this reduce the cost of printing the form, it will also require less storage space when filing completed forms.

Table 2 is an example of how to organize the form for quick examination of an area based on the classification hierarchy. Note that the form reflects the concern of the data processing analyst with reclaimed areas, natural vegetation and activities associated with a strip mining operation. A word or couple of words, check marks or circles are the methods employed to facilitate the completion of this form.

Additional Equipment

The following is a listing and brief explanation, where necessary for clarification, of various items essential or useful in field studies.

Clipboard
Ball point pens
Pencils
Driftline pen and ink -- For annotations on photography and maps.

TABLE 2. MODIFIED GROUND TRUTH FORM.

I. Mine Name _____ Site No. _____

Date _____

Time _____

Crew _____

II. Film Type _____ Filter _____

Roll No. _____ Exposure No. _____

Azimuth _____ Subject _____

III. Site Description (check appropriate site)

Forested Area ( ), Woodland ( ), Savannah ( ), Reclaimed Area ( ),

Mining Activities ( )

Major Plant/Tree Species _____, _____, _____

* Understory vegetation types if area is forested or a woodland

_____, _____, _____, _____

Average Plant Height _____

Ground/Canopy Density: 0-10% ( ) 10-20% ( ) 20-30% ( ) 30-40% ( )
40-50% ( ) 50-60% ( ) 60-70% ( ) 70-80% ( )
80-90% ( ) 90%> ( )

Plant Condition:	Good-Excellent ( )	Fair-Good ( )	Poor ( )
	Vigorous Healthy Above Avg. Growth	Vigorous Healthy Avg. Growth	Chlorotic Wilted

Diseased ( ) Insect-Damaged ( )

Necrosis ↳ % of Site _____

Stage of Growth: Immature ( ) Mature ( ) Flowering ( ) Heading ( )

Soil Type _____ Soil Color _____

Soil Texture _____

Water Regime: Dry ( ) Moist ( ) Wet ( ) Boggy ( )

Slope _____

Aspect _____

Exposed Rocks, Boulders, Etc. Y N Approximate Size _____

Comments: _____

_____

_____

_____

Scale (1/10 inch increments) -- Imagery at a scale of 1:24000 or 1:62500 provides good surface detail on the photography, ease of detail matching to a standard scale map, and allows direct measurements of approximately one inch to a mile. One-tenth inch increments are advantageous in relating vehicle odometer indications to the maps and photographs.

Field identification guides -- Not everyone can hope to be a proficient botonist or forester, at least in respect to making rapid species identifications in the field. If project requirements call for species differentiation, the average field worker will find a good identification guide an invaluable source of information.

Camera and film -- Photographic records are often valuable in later analysis of specific areas or individual training samples.

Plant press -- If field identifications are not possible, a plant press may be used to preserve samples for later identification and specimen for study by future crews.

Transparent tape -- It is often advisable, when making annotations on photography, to cover them with a protective piece of tape. Vinyl transparent tape eliminates glare and readily accepts ink for additional notations.

Collection bags -- Transparent plastic bags are good for collecting plant species. If the samples must be retained for more than 2 or 3 days, a plant press is more practical.

Adhesive labels -- Labels may be used to identify training samples on photography or to mark collected vegetation samples.

Compass -- The importance of precise orientation in the field cannot be overemphasized. Not only is it embarrassing to be lost, but one may be hard pressed to locate a number of training samples studies under such a condition.

REFERENCES

1. Wenderoth, S. and E. Yost. Multispectral Photography for Earth Resources, Long Island University, Greenvale, New York, pp. 4.1-4.17, 1972.

2. Packer, Paul E. Rehabilitation Potential and Limitations of Surface-Mined Land in the Northern Great Plains, U.S. Department of Agriculture, Technical Bulletin Number , pp. 4-7, 1974.

3. Trewartha, Glenn T., A. H. Robinson, and E. H. Hammons. Fundamentals of Physical Geography, McGraw Hill Book Company, pp. 322-325, 1961.

4. Fowells, H. A. Silvics of Forest Trees of the United States, U.S. Department of Agriculture, Handbook Number 271, pp. 411-422, 1965.

5. Whitley, Sidney L. Low Cost Data Analysis Systems for Processing Multispectral Scanner Data, NASA-JSC Earth Resources Laboratory at NSTL, Report Number 157, January, 1977.